

С.Е. Вадяев, С.И. Дубинин *
 Самарский государственный университет

**Параллельный корпус текстов
 и современная стилометрия**
 (на материале электронного корпуса «DER»)

S. Vadjaev, S. Dubinin
Parallelkorpus und moderne Stilometrie
 (am Beispiel des elektronischen „DER“-Korpus)

Im Aufsatz werden praktische Aspekte des Aufbaus und des Forschungsgebrauchs eines deutsch-russischen EDV-basierten Parallelkorpus der öffentlichen UNO-Texte behandelt.

Одной из интенсивно формирующихся областей современного языкознания является корпусная лингвистика, занимающаяся разработкой принципов построения лингвистических корпусов данных с использованием компьютерных технологий. Несмотря на ее бурное развитие и активную работу по созданию лингвистических корпусов, доля многоязычных корпусов текстов по сравнению с одноязычными остается малой. Отставание многоязычных ресурсов вызвано сложностью реализации таких корпусов, что тормозит развитие контрастивной лингвистики, многоязычной лексикографии, исследований по переводоведению. Большую ценность представляют *параллельные корпуса текстов* (ПКТ), т. е. текстовые массивы, включающие тексты на одном языке и их переводы. Использование ПКТ помимо преимуществ одноязычного корпуса при изучении отдельного языка предоставляет возможности для проведения контрастивных исследований, анализа передачи различных языковых явлений в разных языках, поиска использующихся в переводческой практике

* © Вадяев С.Е., Дубинин С.И., 2005.

эквивалентов, создания переводных словарей и обучения иностранному языку. ПКТ представляет интерес и как материал для исследования языка переводных текстов и их отличий от языка текстов, изначально написанных на данном языке.

Использование электронного корпуса текстов позволяет эффективнее и шире применять квантитативные (количественные) методы, в частности для выявления стилистических особенностей входящих в его состав текстов. Применение этих методов является неотъемлемой частью при составлении крупных корпусов текстов и при их анализе. Их целью является исследование всех квантифицируемых (поддающихся количественной оценке) свойств текста для их последующего описания, сравнения друг с другом, классификации по различным параметрам и т. д. Полученные данные подвергаются анализу, на основе которого выявляются закономерности употребления отдельных языковых единиц. При этом особое внимание уделяется установлению частотности и дистрибуции отдельных языковых единиц или их групп, стохастических зависимостей между ними, соотношений между внутренними характеристиками текста и внешними факторами. Таким образом возможно гипотезирование закономерностей, определяющих стилистические особенности текстов, выявление специфических характеристик отдельных типов текстов и разграничение самих стилей.

Данная статья посвящена комплексному решению лингвистических проблем, связанных с построением немецко-русского ПКТ и его использованием для выявления особенностей текстов официально-делового стиля современного немецкого языка.

Для современной стилистики характерна тенденция к дескриптивизации, а количественные показатели служат важным источником данных о текстах. На том факте, что различия между стилями и жанрами носят преимущественно статистический характер, основана *статистическая стилистика*. Э.Г.Ризель и Е.И.Шендельс, говоря о стиле как о «совокупности лексических, грамматических, фонетических выразительных и стилистических средств, используемых в определенных функциональных сферах для выражения определенного содержания», подчеркивают, что «нельзя ограничиваться лишь перечислением отдельных элементов этой системы», поскольку необходимо также давать «квалитативное и квантитативное описание»

их употребления [1, 49–51]. В этой связи отмечаются преимущества комбинирования двух этих основных методов стилистического анализа. Такое сочетание позволяет объединить достоинства этих методов и нивелировать их недостатки.

В качестве основы для такого исследования служит корпус — *репрезентативная выборка* текстов, составленная с учетом таких факторов, как функционально-стилистические особенности, письменная или устная форма существования, содержание. Объем выборки должен быть достаточно большим для того, чтобы достоверно определить закономерности, отделив их от случайных, нетипичных случаев употребления. Количественные данные, полученные на основе такой выборки, служат «симптомами» качественных характеристик и представляют основу для последующей интерпретации. Ограниченное количество ПКТ вообще, в особенности для пары языков немецкий / русский, недостаточная разработка концепции их составления и обработки определяют актуальность данных исследований [9, 104–120].

В рамках совместного проекта *кафедры немецкой филологии Самарского госуниверситета и института немецкого языкознания университета г. Вюрцбург (ФРГ)* начата работа по созданию немецко-русского ПКТ. Проект назван «*DER-Korpus*» (Deutsch-Russisches Parallelkorpus) и представляет собой двуязычный немецко-русский ПКТ (см.: <http://germanistik.ssu.samara.ru>). Предлагаемая статья является первым итогом работы по созданию и использованию ПКТ «DER». Была описана лингвистическая концепция и разработана технология построения немецко-русского ПКТ, наглядно представлены возможности использования субкорпуса ПКТ для исследований в области стилистики. В рамках работы детально апробирована методика автоматической стыковки параллельных текстов на немецком и русском языках, что не служило ранее предметом рассмотрения лингвистов, а также осуществлено практическое приложение немецкого субкорпуса ПКТ «DER» для выявления закономерностей употребления модальных глаголов в текстах официально-делового стиля современного немецкого языка [2].

Значимость исследования состоит в разработке теоретической базы для моделирования и реализации ПКТ, а также в обосновании возможностей применения ПКТ «DER» для анализа стилистических характеристик конкретных типов текстов. Данная работа касается

также круга проблем, затрагивающих особенности официально-делового стиля в современном немецком языке, места документов ООН, как самостоятельного типа текстов, в динамичной системе функциональных стилей современного немецкого литературного языка. Несмотря на то, что официально-деловой стиль является многоплановым феноменом и в спектре политических типов текстов проявляет в последние десятилетия значительную динамику, связанную с расширением международной функции немецкого языка, за последнее время не появилось большого количества работ, посвященных данной тематике.

Материалом для исследования послужил массив электронных документов ООН (www.un.org) на немецком и русском языках. Основу корпуса составляют шесть докладов о работе ООН – ежегодных выступлений Генерального Секретаря ООН перед Генеральной Ассамблеей в период 1997–2000 гг. Общий объем текстов корпуса составил около 900 страниц и включает примерно 300 000 словоупотреблений (на каждый из языков корпуса 150 000).

Цель исследования заключалась в разработке *технологии составления немецко-русского ПКТ* и определении *релевантных возможностей его применения для выявления стилистических особенностей текстов официально-делового стиля*. На первом этапе ставилась задача разработать технологию автоматической обработки текстов для ПКТ, средств их представления и анализа. На втором этапе на основе данных количественного и качественного стилистического анализа немецкого субкорпуса ПКТ «DER» выявлялись *закономерности частотно-распределительных характеристик немецких модальных глаголов и их зависимость от общей коммуникативно-функциональной направленности* представленного в ПКТ «DER» типа текстов (ежегодные доклады Генерального Секретаря ООН).

ПКТ «DER» может служить в качестве практической основы и инструментария для дальнейших стилистических, лексикологических, грамматических и контрастивных исследований, позволяя:

проводить сравнительный анализ особенностей немецкого и русского языков в плане лексического состава, морфологии и синтаксиса (статистический и сравнительный анализ лексических структур и синтаксических конструкций в параллельных текстах);

— осуществлять поиск эквивалентов для слов и коллокаций, примеров употребления лексических единиц, их перевода на другой язык в зависимости от контекста употребления;

— проводить исследования в области стилистики (рассмотрение способов передачи стилистических особенностей в текстах на разных языках, определение приемов и стилистических средств, при помощи которых в языке перевода передается содержание оригинального текста);

— разрабатывать программные средства для проведения стыковки текстов, автоматического поиска и извлечения эквивалентов для отдельных лексем и коллокаций;

— исследовать явления в одном из языков корпуса.

Использование машинных корпусов текстов с 1960-х гг. открыло новые возможности для проведения лингвистических исследований. Современные технологии позволяют существенно увеличить объем обрабатываемых данных, повысить эффективность их обработки. Сам по себе корпус текстов не несет новой информации о языке, являясь «копилкой», куда складываются тексты, язык в своем естественном употреблении. При помощи компьютера можно эффективно управлять содержащейся в корпусе информацией, ее представлением для пользователя. Именно это новое представление языковых данных дает возможность получить новую информацию о языке, являясь базисом для дальнейшей интерпретации. Но полученные на основании корпуса данные действительны в первую очередь для используемого корпуса и не могут быть без оговорок обобщены для всего языка.

Термин «*корпус*» является неоднозначным, став собирательным и обозначая в современной лингвистике самые разнообразные объекты. Он нередко утрачивает строгость в употреблении (ср. «корпус высказываний / примеров употребления»), став «модным». Общим является то, что во всех случаях речь идет о некой *подборке языкового материала*, и, как правило, имеется в виду подборка полных текстов или их достаточно крупных фрагментов [3, 14-15]. Нет единства мнений и по поводу того, является ли корпусом выборка изолированных предложений или отдельных слов. Для наглядности рекомендуется использовать термин *корпус текстов*, который употребляется наравне с термином *корпус*, обозначая подборки текстов или их отдельных частей. Под текстом нами понимается относительно самостоятельная, содержательно связанная последовательность естественных языковых высказываний, а под частью текста — его связанный,

сплошной фрагмент. Корпус текстов следует отличать от подборки примеров, таких как набор лексикографических карточек, поскольку в данном случае из используемых текстов относительно произвольно извлекаются отдельные цитаты и условие содержательной когерентности не выполняется.

Нет единства мнений также по поводу того, должен ли корпус содержать только тексты, которые являются результатом спонтанных, естественных ситуаций коммуникации, или данный термин применим также к текстам, искусственно возникшим вне актов языкового общения. Поскольку в дискуссии об эмпирических данных, собираемых для описания языка, центральную роль играет противопоставление корпуса и интуиции, неуместно обозначать термином «корпус» набор выдуманных (возникших вне акта коммуникации) примеров. Высказывания, возникшие в результате коммуникативного намерения, нельзя приравнивать к примерам, которые искусственно создает считающий себя «компетентным» носитель языка. Неоднозначными являются также данные, полученные в результате неспонтанного акта коммуникации.

Опираясь на К. Бауша [4, 78] и Т. Бунгартена [5, 34], мы придерживаемся следующего определения: корпусом является ограниченное количество языкового материала, состоящего исключительно из (письменных или устных) текстов или их частей.

Значительными проблемами остаются формулировка принципов составления ПКТ, описание исходных для них массивов текстов, автоматическое определение границ структурных единиц текста (абзацев и предложений) и определение подхода к автоматической стыковке параллельных текстов. Кроме того, необходимо ориентироваться в программном обеспечении для корпусов текстов и в программной оболочке, разработанной для представления текстов ПКТ и для осуществления поиска.

При составлении ПКТ, в отличие от одноязычных и сравнительных корпусов текстов, следует учитывать и фактор межкультурных связей. Множество текстов исходного языка, пригодных для включения в ПКТ, составляют лишь те тексты, которые были переведены на второй язык. Так, подбор текстов для сбалансированного русско-немецкого ПКТ является сложной задачей. Это связано с тем, что

не всегда возможно найти тексты, аналогичные по тематике, жанру и относящиеся к одному периоду времени. Хотя для данной пары языков существует большое количество переводов, различия в культурных традициях приводят к сильному дисбалансу в структуре субкорпусов. Таким образом, представляется проблематичным получить «зеркальные» субкорпуса оригинальных и переводных текстов.

Кроме этого при подборе текстов для ПКТ приходится руководствоваться не только теоретическими соображениями, но и учитывать ряд практических факторов, влияющих на возможности использования тех или иных текстов в ПКТ [10, 138-143]. Практически все современные тексты (особенно художественные произведения) защищены авторскими правами, что исключает их перевод в электронную форму и использование без разрешения правообладателя. Получения прав для использования электронной версии текста (даже для научных целей) представляет собой значительную проблему. Поэтому для ПКТ «DER» было решено использовать исключительно тексты, которые не подпадают под действие авторского права и могут быть использованы без ограничений.

ПКТ «DER» изначально разрабатывался как пополняемый корпус параллельных текстов, а его объем не ограничивался определенным набором текстов заданной тематики. Это значит, что он может быть расширен за счет других параллельных текстов, а в зависимости от специфики данных текстов может варьироваться и область применения корпуса. Тексты, добавляемые в корпус, могут образовывать *субкорпуса*, включающие в себя тексты определенного типа. Такие субкорпуса могут предназначаться для решения специализированных задач, рассматриваться отдельно от остальной части корпуса или сравниваться со всем корпусом для выявления особенностей текстов, входящих в них.

Но важнейшей проблемой при составлении ПКТ «DER» остается стыковка параллельных текстов (*alignment*). Решение данной задачи было разделено на два этапа: выявления границ структурных элементов текста и определение параллельных сегментов. Основой для стыковки параллельных текстов служит структурное сходство между оригиналом и переводом, проявляющееся на разных уровнях текста в различной степени (см. рис. 1).

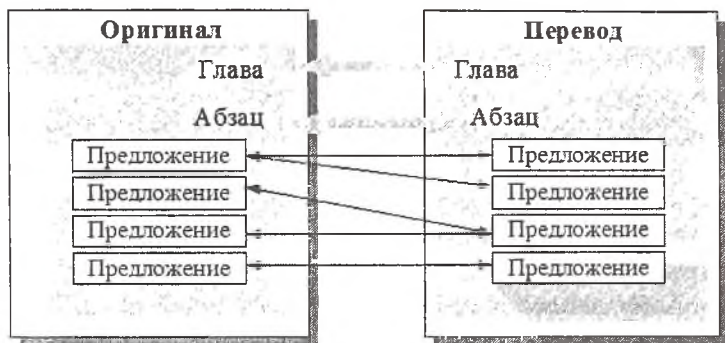


Рисунок 1
Общий процесс стыковки текстов в ПКТ «DER»

На момент создания ПКТ «DER» не было найдено спецпрограмм, предназначенных для выполнения данной операции с размеченными (таггированными) текстами на немецком и русском языках. По этой причине для стыковки текстов была создана специализированная программа, учитывающая особенности текстов корпуса и специально приспособленная для работы с немецкими и русскими текстами. Для определения параллельных сегментов программа использует статистико-стохастический подход, что обеспечивает высокую надежность результатов стыковки при минимальных временных затратах на ее разработку. При стыковке параллельных текстов сначала стыкуются крупные сегменты текста, а затем происходит проверка правильности и начинается обработка малых сегментов на следующем уровне. Такой подход позволяет минимально сократить количество ошибок в процессе автоматической стыковки.

Для наглядного представления текстов ПКТ «DER» и осуществления эффективного поиска в них была разработана специальная программная оболочка. При этом особое внимание уделялось ее функциональности: программа предназначена для работы с таггированными параллельными текстами с лингвистической разметкой в формате XML (Extensible Markup Language), соответствующей международному стандарту TEI (Text Encoding Initiative) - TEI Lite, версия для XML (см.: www.tei-c.org). Стандарт TEI-разметки представля-

ет собой набор предложений по кодировке электронных текстов различных типов для лингвистических целей. Особое значение при разработке программной оболочки уделялось следующим функциям: эффективный поиск в текстах, возможность различного представления его результатов (параллельный конкорданс или т. н. KWIC-индекс — key word in context), автоматическое составление полного списка словоформ, обеспечение доступа и поиска в текстах корпуса через Интернет (см. рис. 2).

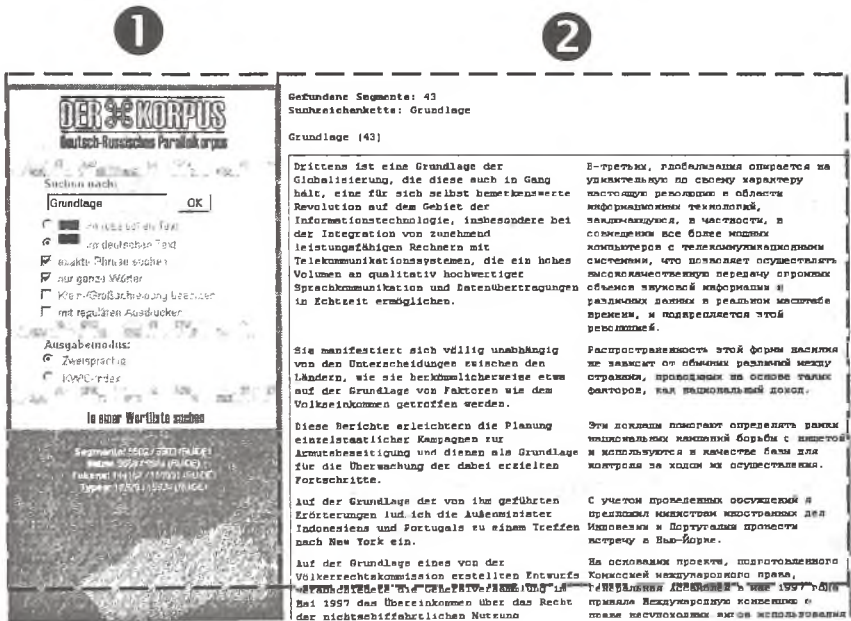


Рисунок 2

Двусторонний интерфейс программной оболочки ПКТ «DER»

Прикладная часть данного исследования включает оценку использования количественных методов в стилистике, определение актуальных проблем исследования текстов официально делового стиля, описание основных характеристик входящих в ПКТ «DER» текстов.

На основе представленных в ПКТ текстов рассматривалась зависимость между общей коммуникативной функцией текста и частотно-распределительными характеристиками немецких модальных глаголов.

Тексты ООН, входящие в ПКТ «DER», представляют особый интерес для изучения особенностей официально-делового функционального стиля в немецком языке и опытного подтверждения предположений, касающихся особенностей таких текстов. Использование корпуса дало возможность быстро и надежно получить результаты обработки большого массива текстов, что обеспечивает получение точных данных. Используя тексты ООН, входящие в ПКТ «DER», можно как выявить языковые средства, характерные для данного типа текста, так и дать подробную картину их употребления в текстах данного типа. Поскольку в настоящее время существует дефицит исследований, изучающих отличительные особенности текстов официального характера, где на основе их репрезентативной подборки проводился бы не только обзор отдельных, пусть даже и типичных для данного функционального стиля явлений, но и комплексный анализ текстов со статической оценкой частотности отдельных явлений, именно такой подход обеспечивает достоверные результаты изучения функционально-коммуникативных особенностей данного типа текстов.

Опираясь на описанные в большинстве практических грамматик немецкого языка значения модальных глаголов [6, 131-136; 7] и на теорию речевых фонов [8, 1882], мы попытались определить, какие значения модальных глаголов являются наиболее употребительными в текстах ПКТ «DER» на немецком языке и соотнести полученные данные с общей коммуникативной интенцией текстов официально-делового функционального стиля (доклады). ПКТ «DER» как нельзя лучше подходит для решения данной задачи, поскольку семантика модальных глаголов в немецком тексте, как правило, наглядно раскрывается и в параллельных текстах на русском языке.

Исследование осуществлялось в три этапа. На первом этапе был произведен поиск всех модальных глаголов в корпусе и определена их частотность в немецких текстах ПКТ «DER» (для всего корпуса и для каждого текста в отдельности). На втором этапе был проведен анализ значения каждого конкретного употребления модального гла-

гола в текстах ПКТ «DER» на немецком языке в соответствии со значения модальных глаголов, описанными в практических грамматиках, и с теорией речевых фонов. На завершающем этапе было проведено сравнение полученной информации об употреблении каждого модального глагола с общей коммуникативно-функциональной направленностью текстов корпуса «DER». Исходя из того, что включенные в ПКТ «DER» тексты ООН являются типичными текстами официально-делового функционального стиля, мы предположили, что прагматическая функция *долженствования* и *директивный характер* данных текстов, должны соответствующим образом отражаться в речевом фоне употребления модальных глаголов.

Чтобы провести анализ употреблений модальных глаголов, необходимо выявить в корпусе все их появления в контексте. Использование ПКТ «DER» позволяет быстро и эффективно выполнять эти сложные поисковые запросы. Чтобы найти все употребления модальных глаголов в корпусе, для каждого отдельного глагола был составлен *поисковый запрос* с использованием регулярных выражений. С его помощью можно осуществлять комплексный поиск форм любого слова. Например, поисковый запрос для глагола *dürfen* выглядел следующим образом:

`\Wdürf(en|t)\W\Wdarf(st)*\W\Wdurft(e|est|en|et)\W\WgedurftW.`

Программная оболочка корпуса обрабатывает запрос и производит поиск всех соответствий, содержащихся в текстах корпуса. Результаты поиска выводятся в удобной и наглядной форме, что облегчает их анализ и обработку. Программа подсчитывает общее количество найденных глаголов, отображает отдельные формы и их количество. Она позволяет не только найти все употребления для модальных глаголов, но и выявить их частотность во всем корпусе и в каждом отдельном тексте корпуса. На первом этапе был произведен автоматический поиск всех употреблений модальных глаголов *dürfen*, *können*, *wollen*, *mögen*, *sollen* и *müssen* в немецких текстах ПКТ «DER». Немецкая часть ПКТ «DER» содержит 153931 словоупотребление. Среди них было найдено 838 употреблений модальных глаголов. Статистика употреблений глаголов представлена на диаграмме 1.

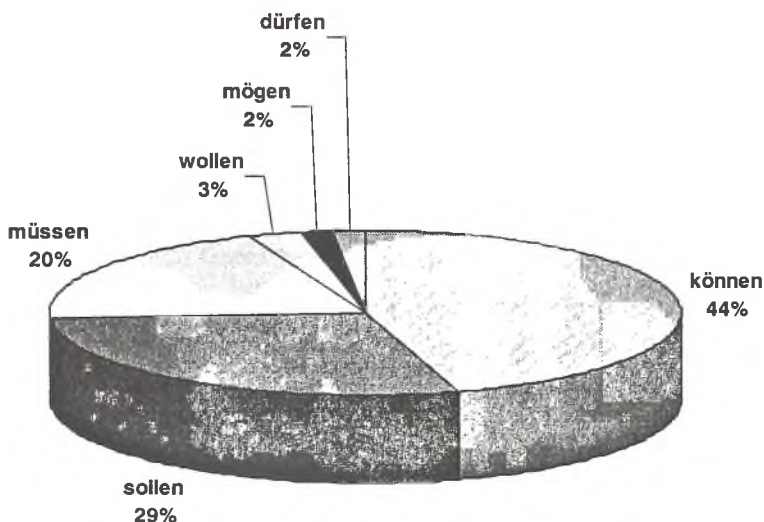


Диаграмма 1

*Распределение модальных глаголов по частотности
в немецких текстах ПКТ «DER»*

Как показывает частотность употребления модальных глаголов, *dürfen*, *mögen* и *wollen* встречаются в представленных текстах значительно реже, чем *können*, *müssen* и *sollen*. Однако показатели частотности для всего корпуса не дают исчерпывающих сведений о частотности каждого из модальных глаголов в отдельно взятом тексте. Можно предположить, что частотность отдельно взятого глагола не зависит от типа текста и, следовательно, может значительно варьироваться в различных экземплярах текста.

Поскольку мы предположили, что частотно-распределительные характеристики модальных глаголов являются параметром, свойственным определенному типу текстов (текстам ООН, относящимся к официально-деловому функциональному стилю), была выполнена их проверка в каждом отдельно взятом тексте ПКТ «DER». Результаты этого анализа представлены в табл. 1. Как видно из данных, частотно-распределительные характеристики наиболее употребитель-

ных глаголов *können*, *müssen* и *sollen* сохраняются в каждом из текстов корпуса. Это означает, что частотность данных глаголов является отличительной чертой представленного в корпусе типа текстов и не варьируется в зависимости от экземпляра текста. По частотно-распределительным характеристикам вторая группа глаголов (*dürfen*, *mögen* и *wollen*) проявляет по сравнению с более частотными глаголами (*können*, *müssen* и *sollen*) заметную вариативность. Это обусловлено тем, что употребление этих глаголов в данном типе текстов намного реже (на эти 3 глагола приходится менее 10% всех употреблений модальных глаголов в текстах корпуса). Низкая частотность данной группы модальных глаголов и значительная вариативность их частотно-распределительных характеристик в отдельных экземплярах текста говорят о том, что глаголы играют второстепенную роль в представленном типе текстов и их употребление зависит от конкретного экземпляра текста.

Таблица 1

*Частотно-распределительные характеристики
модальных глаголов в отдельных текстах ПКТ «DER»*

Модальный глагол	Тексты ПКТ «DER» на немецком языке						Кол-во употребл. (все тексты)
	A-52-1	A-53-1	A-55-1	A-56-1	A-57-1	A-58-1	
können	43	67	64	73	62	72	381
sollen	29	46	50	35	30	50	240
müssen	25	26	37	31	19	28	166
wollen	0	8	6	6	0	3	23
mögen	4	0	3	1	2	6	16
dürfen	1	2	3	3	1	3	13
Всего:	102	149	163	149	114	162	839

Для того чтобы подтвердить, что приведенные выше частотно-распределительные характеристики модальных глаголов составляют особенность представленного в ПКТ «DER» типа текстов, был проведен анализ частотно-распределительных характеристик модальных глаголов в корпусе «Der deutsche Wortschatz» (далее «DDW», см.: www.wortschatz.uni-leipzig.de, проект университета г. Лейпциг/ФРГ). Данный корпус представляет собой сверхбазу данных, включающую в себя тексты различных типов (в основном газетные, а также специализированные тексты). На 2003 г. корпус «DDW» включал 35 млн. предложений, содержащих 500 млн. словоупотреблений. Объем корпуса позволяет авторам проекта говорить о том, он охватывает основную часть современной немецкой письменной речи. В текстах «DDW» нами был произведен поиск всех употреблений модальных глаголов *können*, *dürfen*, *mögen*, *wollen*, *müssen* и *sollen*. Поскольку «DDW» охватывает различные типы текстов широкой тематики и имеет значительный объем, данные о частотно-распределительных характеристиках модальных глаголов, полученные на его основе, могут считаться «масштабом» их употребления в целом. На диаграмме 2 представлены результаты анализа.

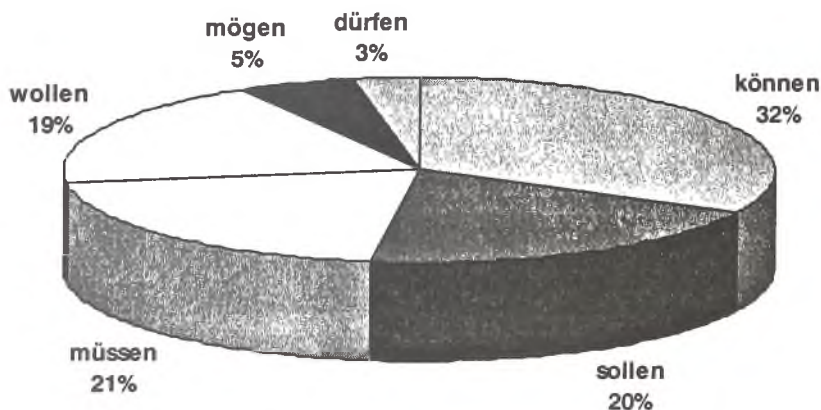


Диаграмма 2

Распределение модальных глаголов по частотности в корпусе «DDW»

Очевидно, что самым употребительным в «DDW» является *können*, что совпадает с данными, полученными на основе ПКТ «DER». Но заметно, что доля употреблений этого глагола в «DDW» меньше, чем в ПКТ «DER» («DER» – 44%, «DDW» – 32%). Более интересны изменения частотно-распределительных характеристик *wollen*, *müssen* и *sollen*. Количество употреблений этих глаголов в корпусе «DDW» практически равно. Особенно значителен, по сравнению с ПКТ «DER», прирост употреблений *wollen* в корпусе «DDW» («DER» – 3%, «DDW» – 19%). Количество употреблений *sollen* в корпусе «DDW» ниже, чем в немецких текстах ПКТ «DER» («DER» – 29%, «DDW» – 20%). Диаграмма 3 отражает различия частотно-распределительных характеристик модальных глаголов в обоих корпусах.

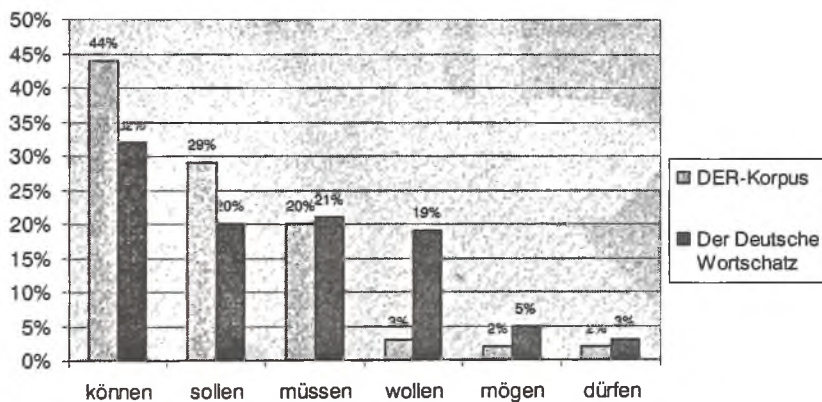


Диаграмма 3

*Распределение модальных глаголов по частотности
в корпусе «DDW» и в немецком субкорпусе ПКТ «DER»*

При сравнении двух корпусов отчетливо видно, что частотно-распределительные характеристики некоторых модальных глаголов в представленных корпусах текстов значительно варьируются. Это можно объяснить тем, корпуса содержат разные типы текстов с отличающейся коммуникативно-функциональной направленностью. Изменения в структуре употребления модальных глаголов являются отражением этих различий. «DDW» можно рассматривать как общий корпус, поскольку он включает в себя тексты различных типов. Пе-

ред таким корпусом ставится задача охватить не какую-то определенную предметную или проблемную область, а дать широкий срез наибольшего количества текстов самого различного вида. «DDW» имеет большой объем и может использоваться как базис для сравнения со специализированными корпусами. Таким специализированным одноязычным корпусом можно считать немецкую часть ПКТ «DER», т. к. он представляет собой подборку текстов определенного типа (официальные тексты ООН, относящиеся официально-деловому функциональному стилю).

Учитывая эти предположения, мы попытались интерпретировать статистические данные, полученные на основе этих двух корпусов. Наиболее очевидным и легко объяснимым с учетом функциональных особенностей текстов корпуса «DER» различием является низкая частотность глагола *wollen* («DER» – 3%, «DDW» – 19%). Это связано с тем, что основным значением данного глагола является выражение желание или намерения. Выражения субъективных желаний и использование интрасубъективно-волитивного речевого фона мало соответствует коммуникативно-функциональной направленности текстов официально-делового стиля. Это также может служить объяснением для более низкой частотности глагола *mögen* в текстах немецкой части ПКТ «DER».

Основной характеристикой глаголов *müssen* и *sollen* является выражение различных форм долженствования. В корпусе «DDW» частотно-распределительные характеристики этих глаголов примерно одинаковы. В немецкой части текстов ПКТ «DER» мы наблюдаем значительное увеличение количества употреблений *sollen*. По нашему мнению, это связано с тем, что в текстах ООН, как правило, долженствование выражается как проявление чужой воли, оно носит предписывающий характер. Именно это определяет значительное увеличение количества употреблений *sollen*.

Последнее различие в употреблении модальных глаголов в корпусе «DDW» и ПКТ «DER» заключается в значительной разнице частотно-распределительных характеристик глагола *können*. Исходя из особенностей функционального стиля представленных в ПКТ «DER» текстов, можно предположить, что прескриптивный характер текстов ООН проявится в более высокой частотности употребления глагола *dürfen* и некотором снижении частотности *können*. Глагол

dürfen в основном выражает значение разрешения, которое более соответствует интенции официально-делового функционального стиля. Тем не менее, сравнивая частотно-распределительные характеристики *können* и *dürfen*, мы наблюдали противоположное явление. По сравнению с корпусом «DDW» количество употреблений в немецких текстах ПКТ «DER» заметно выше (DER» — 44%, «DDW» — 32%), в то время как количество употреблений модального глагола *dürfen* незначительно снижается («DER» — 2%, а «DDW» — 3%).

Очевидно, что представленные феномены во многом подтверждают предположения о том, что изменения частотно-распределительных характеристик модальных глаголов во многом обусловлены особенностями функционально-стилевой принадлежности представленных в корпусе текстов. Однако изменения частотности отдельных глаголов (*können* и *dürfen*) не совсем совпали с нашими первоначальными предположениями. Для того чтобы интерпретация представленных выше статистических данных была полной, был проведен анализ всех употреблений модальных глаголов в немецких текстах ПКТ «DER».

Несмотря на то, что прагматическая функция долженствования играет важную роль в документах ООН (особенно в докладах Генсека), ее выражение носит менее категоричный, рекомендательный характер. Это косвенно подтверждается преобладанием количества употреблений глагола *sollen* над количеством употреблений *müssen*. Статус ООН и язык дипломатических документов обуславливают общую тенденцию снижения категоричности выраженного модальными глаголами долженствования, при одновременном сохранении коммуникативной функции *долженствования* и *предписывающего, нормативного характера* данных текстов. Большинство употреблений глагола *können* относится к телеологическому речевому фону и служит для ориентации высказывания на определенную цель, а также для выражения возможности ее достижения, что вполне соотносится с общей интенцией представленных в ПКТ «DER» текстов ООН. Волитивный речевой фон играет в текстах ООН скорее второстепенную роль, т. к. он не соответствует общей функционально-коммуникативной направленности официально-делового стиля. Этим фактором обусловлена довольно низкая частотность глаголов *wollen* и *mögen*.

Каждый из модальных глаголов реализует целый спектр модальных значений, но в каждом конкретном употреблении актуализируется одно из этих значений. При этом значение модального глагола обуславливается речевым фоном высказывания. Анализ употреблений модальных глаголов в немецких текстах ПКТ «DER» показывает, что различные значения модальных глаголов отчетливо проявляются в текстах на русском языке, что может служить в качестве дополнительного дифференцирующего признака, отражающего речевой фон данного употребления и позволяющего оценить его зависимости от общей интенции текстов, представленных в ПКТ «DER».

Данное исследование позволяют сделать следующие выводы.

1. ПКТ является универсальным инновационным средством для проведения лингвистических исследований и может служить не только для сопоставительных исследований нескольких языков или поиска переводных эквивалентов, но и в качестве базиса для исследований каждого из языков корпуса (в рамках субкорпусов).

2. Использование комплекса современных технологий (разметки TEI, формата XML, специального программного обеспечения для стыковки текстов и их репрезентации) позволяет оптимизировать процесс создания ПКТ и повысить эффективность его обработки.

3. Тщательная выборка текстов позволяет уже на основе достаточно **компактного** корпуса (принцип оптимальности) получать достоверные данные для всех текстов, отвечающих параметрам отбора.

4. Документы ООН занимают особое место среди политических типов текстов в составе официально-делового стиля современного немецкого языка, что обусловлено их внешнеполитическим характером, статусом их эмитента, усиливающейся интернационализацией немецкого языка.

5. Между общей коммуникативно-функциональной направленностью текста и основными средствами выражения модальности существует зависимость, выражающаяся для модальных глаголов немецкого языка в изменении их частотно-распределительных характеристик, учитывая функционально-стилевую принадлежность текста.

6. Речевой фон употребления модальных глаголов немецкого языка тесно связан с общей коммуникативно-функциональной направленностью текстов определенного типа.

Библиографический список

1. Ризель, Э.Г. Стилистика немецкого языка [Текст]/ Э.Г. Ризель, Е.И. Шендельс. — М., 1975.
2. Вадяев, С.Е. Лингвистические принципы построения и использование корпуса текстов для исследования официально-делового стиля современного немецкого языка (на материале электронного корпуса «DER») [Текст]: автореф. дис. ... канд. филол. наук / С.Е. Вадяев. — Нижний Новгород, 2005.
3. Андрющенко, В.М. Автоматическая обработка текста [Текст] / В.М. Андрющенко // Лингвистический энциклопедический словарь / ред. В.Н. Ярцева. — М., 1990.
4. Bausch, K.-H. Intuition und Datenerhebung in der Linguistik [Текст] // Empirische Textwissenschaft: Aufbau und Auswertung von Text-Corpora. — Königstein, 1979.
5. Bungarten, T. Das Korpus als empirische Grundlage in der Linguistik und Literaturwissenschaft // Empirische Textwissenschaft: Aufbau und Auswertung von Text-Corpora. — Königstein, 1979.
6. Helbig, G., Buscha J. Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht. — Leipzig, 1984.
7. Шендельс, Е.И. Практическая грамматика немецкого языка [Текст] / Е.И. Шендельс. — М., 1988.
8. Zifonun, G., Hoffmann L., Strecker B. Grammatik der deutschen Sprache. — Berlin, 1997. — Bd. 3.
9. Вадяев, С.Е. Использование корпусов параллельных текстов в прикладной германистике [Текст] / С.Е. Вадяев // Слово — Высказывание — Дискурс. — Самара, 2004.
10. Дубинин, С.И. Принципы составления немецко-русского корпуса текстов [Текст] / С.И. Дубинин, Ю.Б. Смольская // Аспекты становления и функционирования западногерманских языков. — Самара, 2003.